**NIST BD joint subgroup meeting minutes for Dec 9, 2013**

Prepared by Sanjay Mishra, Verizon

**Scope**

* Recap Current Deliverables and Share Future Activities

Current Deliverables:

* + V1.0 Documents for Requirements and Sec. & Privacy – for technical editing
  + V1.0 Documents for Def. & Tax., Ref. Arch., and Tech. Roadmap – need harmonization
  + Schedule: every Monday, 1:00PM – 4:00PM EDT till Dec. 23, 2013
  + Goal: finalize all V1.0 draft documents (ready for technical editing) **by Dec. 23, 2013**
  + Reflector: please use [bigdatadeftax@nist.gov](mailto:bigdatadeftax@nist.gov), [bigdataarch@nist.gov](mailto:bigdataarch@nist.gov), and/or [bigdataroadmap@nist.gov](mailto:bigdataroadmap@nist.gov) for discussion

**Monday, December 09, 2013**

1. **Review Action Items**
2. Nancy and Orit to continue to collaborate and update section 3 of RA incorporating from Definition and Taxonomy. Specific updates to include review of Levels 1 and 2 (subsequently Levels 3 and 4) and Framework Provider and System Management
3. System Management needs (section 4.1) updated (Owner: Felix Njeh)
4. Need a synch up with Security (Owner: Orit)
5. All participants should review mind-map to see what is missing, from actors, and activities perspective. (Owner(s): All) For accessing mind-map, please send email to Wo.
6. Security and Privacy – Smaller sub-teams are working on document updates and a version is expected to be available in two weeks
7. **Big Data World Congress:** Wo presented NBD-PWG work and future standardization activities to 2013 Big Data World Congress, Munich, Germany
   1. Wo informed this study group will meet 3 times over the 12 month period. Once in US, followed by Europe and China. Wo is working on securing hosting sites and will likely include University settings.
      1. The first meeting will be in San Diego. The duration is a 4-day form. The meeting may move from instead of starting Jan 28 to some time in Feb. (need to give a 60 day notice for ISO members)
      2. Zurich University and Copenhagen University have offered to host the European meeting
      3. The third meeting will likely be in June in China.
   2. Wo has uploaded presentation (<http://bigdatawg.nist.gov/_uploadfiles/M0270_v1_9179221138.pdf>) he gave at the 2013 BD World Congress in Munich, Germany.
   3. There is no fee for working on the Study Group
   4. Additional details to be published by Wo
   5. Anyone interested in collaborating on this, please reach out to Wo (wchang@nist.gov)

3. **Next meeting: Dec. 16, 1:00PM - 4:00PM EST**

**Definition and Taxonomy**

1. **Definition** 
   * No additional input
2. **Taxonomy**
   * Major objective is to align Taxonomy with Reference Architecture for the first two levels (actors and activities) of the Mind Map for Big Data Application Provider and Big Data Framework Provider.
   * At this meeting, in the best interest of both Def&Tax and RA, Taxonomy will have its own document so that it can continue to expand in-depth, more specifically for components and subcomponents with any affect to the RA’s high-level activities. For V1.0, Taxonomy will provide first two levels of the Mind Map text to RA.
   * Nancy, David, and Keith have contributed some more subcomponents into the Big Data Framework Provider for Platforms and Processing Frameworks.

**Reference Architecture**

Changes to the document

1. Taxonomy Mind Map and RA document are not aligned but after much discussion and with the help from Wo, a consensus was reached for the first two levels of the Mind Map for Taxonomy and RA. Now, Section 4 of RA is in total agreement with the Mind Map for the first two levels – actors and activities.
2. Contribution text for Infrastructure and Physical & Virtual Resources will be provided by **David**.
3. Contribution text for Processing Frameworks will be provided by **Nancy**.
4. Contribution text for Platforms will be provided by **Nancy**.
5. Section 5: Felix will expands Section 5.1 for System Management since managing big data computing cluster is different from traditional client/server practices (**awaiting changes from Felix**)
6. **Orit** will incorporate all contribution text and produce a revise version of N0226 RA document before next meeting.

**Security & Privacy (Arnab Roy)**

1. Arnab reported work in progress and will be presenting soon
   1. Smaller teams are working currently and are at different level of maturity. Arnab expects a version to be available in about 2 weeks.
2. Objective is to link S&P with Reference Architecture

**Technology Roadmap**

1. With the agreement of Taxonomy and RA, the next step is to try to get the Technology Roadmap aligned with the rest of the subgroups’ documents. Next meeting will go over the TR and Wo will inform Carl and the TR subgroup to participate the joint meeting.
2. Pw will continue to review and provide feedback on Section 7.
3. David will ping the TR authors for sections that are independent from subgroups and make report for next meeting.

**Action Items**

1. Orit will compile the following contribution text and revise the M0266 RA document:
   1. Write up text for Infrastructure and Physical & Virtual Resources – **David**
   2. Write up text for Processing Framework – **Nancy**
   3. Write up text for Platforms – **Nancy**
   4. Write up text for System Management – **Felix**
   5. Others…
2. **Arnab** and his subgroup will provide an update version of the SnP in about 2 weeks.
3. **Wo** will notify Carl (Technology Roadmap Co-Chair) and his team to participate the joint meeting to sync-up the TR document with the rest of subgroups
4. **Pw** will provide feedback on TR Section 7.

**Online Attendee List**

1. Wo Chang
2. Nancy Grady
3. Sanjay Mishra
4. David Boyd
5. Orit Levin
6. Arnab Roy
7. Keith Hare
8. Bill Mandrick
9. William Vorhies
10. Ashok Malhotra
11. Brand Niemann
12. Bob Marcus
13. Ian G
14. William Miller
15. Rick Jones
16. Pavithra Kenjige

**Online Chat log**

(12:53 PM) Wo (guest) joined.

(12:59 PM) Nancy Grady (SAIC) joined.

(1:00 PM) David Boyd (Data Tactics) joined.

(1:01 PM) Wo Chang (Host, NIST): Let's wait few minutes.

(1:03 PM) Rick Jones (Joseki Group LLC) joined.

(1:04 PM) Brand Niemann joined.

(1:05 PM) Ashok Malhotra (Oracle) joined.

(1:07 PM) Orit Levin (Microsoft) joined.

(1:08 PM) David Boyd (Data Tactics): To me transaction support is a sub-component of Query Interface

(1:09 PM) William Vorhies joined.

(1:11 PM) Arnab Roy (Fujitsu) joined.

(1:21 PM) Bill Mandrick  joined.

(1:25 PM) Brand Niemann: Here is a Big Data 3Vs and 1C taxonomy I find practically useful: <http://semanticommunity.info/1105_Media_Government_IT_Forum#Slide_12_Extreme_Information>

(1:26 PM) Brand Niemann: Just the text: Volume– “Big Data” = Hadoop MapReduce, Cloudera Impala– Large “In Database” = Teradata, Exadata, Vertica, Greenplum…..Variety– Unstructured Data: Attivio, Clarabridge…Velocity– TIBCO e.g. Enterprise Service Bus -> Business Events -> Active Spaces…Complexity– R, TERR, S+, SAS, Matlab– Complex Event Processing: Business Events

(1:27 PM) William Vorhies disconnected.

(1:29 PM) Bob Marcus joined.

(1:37 PM) Keith Hare, JCC Consulting, Inc. joined.

(1:40 PM) Bill Mandrick : I think that the "Image" part of the taxonomy conflates different types/cetegories.  First, there is ImageFileType (jpeg, gif, tiff, etc.), RenderedImage (drawings), then there is the Content of the Image (PersonImage, FacilityImage, GeospatialObjectImage, etc.).  Then there is how the image is captured/rendered (e.g. Aerial, Ground, Satellite, etc.)  All of this needs to be kept straight.

(1:42 PM) Nancy Grady (SAIC): I agree with @Bill, that there is a lot being glossed over in the data types. However, we're not here to create a data type taxonomy, so some high-level examples in this level four set should be sufficient

(1:42 PM) Bob Marcus disconnected.

(1:48 PM) Bill Mandrick : I understand Nancy's point.  I have already started to work out an Image Ontology/Taxonomy and a DataType Taxonomy that keeps all of this straight.  I will dust it off and post it.  I'd be overjoyed if it helped shape some of the thinking and architecture.

(1:50 PM) Bill Mandrick  disconnected.

(1:54 PM) Bill Mandrick  joined.

(1:54 PM) William Miller joined.

(1:55 PM) William Miller disconnected.

(1:55 PM) Bob Marcus joined.

(1:56 PM) William Miller joined.

(2:05 PM) Sanjay Mishra(Verizon) joined.

(2:05 PM) Brand Niemann disconnected.

(2:06 PM) Quyen Nguyen joined.

(2:10 PM) Rick Jones (Joseki Group LLC): In the software development projects I've been involved with, the numbers to the left of the decimal point indicated planned changes, and those to the right indicated responses to bugs, users, etc.  No reason why we couldn't have a similar schema.

(2:11 PM) Rick Jones (Joseki Group LLC): At the least the RA and Taxonomy should be joined via a glossary.

(2:13 PM) Ashok Malhotra (Oracle) disconnected.

(2:22 PM) William Miller: processing framwork is very important

(2:22 PM) Keith Hare, JCC Consulting, Inc.: How about if we put the networking implementation under infrastructure?

(2:22 PM) William Miller: software is one thing but the right hardware seems to be even more imporant

(2:24 PM) William Miller: we must look at the detail of perforamnce for storage system, cluster, implemtnations, data analytic, and preparation of data

(2:25 PM) William Miller: there re some choice to be made to gain the level of perforamnce needed

(2:25 PM) William Miller: clusters will depend alot on the requriements

(2:29 PM) Rick Jones (Joseki Group LLC) disconnected.

(2:31 PM) PavithraKenjige joined.

(2:32 PM) William Miller disconnected.

(2:32 PM) Nancy Grady (SAIC): @Rick, perhaps we should just label the taxonomy document as beta without a number (or beta 1) so people know the details are not complete. Having Levels 1 and 2 OK means the definitions and reference architecture version 1 documents can be completed.

(2:37 PM) William Miller joined.

(2:41 PM) William Miller: transport of various formats should be done via a data compression format whcih will save space and latency but most important it make the data format agnostic

(2:41 PM) William Miller: the formtat would related to metadata about the data that is requirested the file type is a comrpession format

(2:43 PM) William Miller: BZIP2 is used with Hadoop for example

(2:43 PM) William Miller: ZIP is generally used with many other systems

(2:46 PM) William Miller disconnected.

(2:48 PM) Ian G, CMU SEI joined.

(2:59 PM) Bob Marcus: es

(3:01 PM) Bob Marcus: I have been on the Web conference but not the telephone line.

(3:02 PM) Bob Marcus: I think that we need some type of guidelines on how to use the deliverables for differn classes of users e.g ISO, end-users, vendors, government agencies etc.

(3:10 PM) Quyen Nguyen disconnected.

(3:10 PM) Orit Levin (Microsoft) disconnected.

(3:10 PM) David Boyd (Data Tactics) disconnected.

(3:10 PM) Keith Hare, JCC Consulting, Inc. disconnected.

(3:10 PM) Bob Marcus disconnected.

(3:10 PM) Ian G, CMU SEI disconnected.

(3:10 PM) PavithraKenjige disconnected.